ORIGINAL

**NATIONAL UNIVERSITY OF SINGAPORE**

**EXAMINATION FOR**

(Semester II: 2017/2018)

**EE5907 – PATTERN RECOGNITION**

MAY 2018 – Time Allowed: 2.5 Hours

---

INSTRUCTIONS TO CANDIDATES

1. This paper contains **FOUR (4)** questions and comprises **FIVE (5)** printed pages.
2. All questions are compulsory. Answer **ALL** questions.
3. The total mark is **ONE HUNDRED (100)**.
4. This is a **CLOSED BOOK** examination. One A4-size formula sheet is allowed.
5. Programmable calculators are not allowed.

Q1 (25 marks). Subquestions (a) and (b) can be answered independently.

(a) Consider training data $x_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $x_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $x_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $x_4 = \begin{pmatrix} -1.5 \\ 0 \end{pmatrix}$ with corresponding class labels $y_1 = 1, y_2 = 1, y_3 = 0, y_4 = 0$. What is the 3-NN estimate of the class label posterior probabilities of data point $x_5 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, where the distance metric used is the Euclidean distance? What is the MAP classification of data points $x_5$? Repeat the above with the Manhattan distance metric.

(6 marks)

(b) Consider a 2-class naïve Bayes classifier with one binary feature and one Gaussian feature. More specifically, class label $y$ follows a categorical distribution parametrized by $\pi$, i.e., $p(y = c) = \pi_c$. The first feature $x_1$ is binary and follows a Bernoulli distribution: $p(x_1|y = c) = \text{Bernoulli}(x_1|\theta_c)$. The second feature $x_2$ is univariate Gaussian: $p(x_2|y = c) = \mathcal{N}(x_2|\mu_c, \sigma_c^2)$. Let $\pi = [0.7\ 0.3]$, $\theta = [0.2\ 0.6]$, $\mu = [-1\ 1]$ and $\sigma^2 = [1\ 1]$.

(i) Compute $p(y|x_1 = 0)$. Note that result is a vector of length 2 that sums to 1.

(6 marks)

(ii) Compute $p(y|x_2 = 1)$. Note that result is a vector of length 2 that sums to 1.

(7 marks)

(iii) Compute $p(y|x_1 = 0, x_2 = 1)$. Note that result is a vector of length 2 that sums to 1.

(6 marks)

Q2 (25 marks). Subquestions (a) and (b) can be answered independently.

(a) Consider a Poisson distribution $p(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$. Suppose we observe $N$ independent samples from the Poisson distribution: $D = \{x_1, \cdots, x_N\}$.

(i) What is the maximum likelihood (ML) estimate of $\lambda$? Show your steps to get full credit.

(7 marks)

(ii) Suppose we use ML estimate of $\lambda$ to predict new data $x_{N+1}$. What problems might arise? Describe a solution to avoid this problem.

(2 marks)

(b) Consider the same distribution and data from part (a). Assume the conjugate prior distribution $p(\lambda) = \text{Gamma}(\lambda; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\lambda}$, where $\Gamma(\cdot)$ is the Gamma function (not to be confused with the Gamma distribution). You may or may not find the following identities useful: $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$, and $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$.

(i) The posterior distribution $p(\lambda|D)$ is also a Gamma distribution with parameters $\alpha', \beta'$. What are $\alpha'$ and $\beta'$? Show your steps.

(7 marks)

(ii) What is the posterior predictive distribution $p(x_{N+1}|D)$? Show your steps. Your final expression can contain the Gamma function.

(9 marks)

Q3 (25 marks). Subquestions (a), (b) and (c) can be answered independently.

(a) Given five data points:

$$x_1 = \begin{pmatrix}0\\0\end{pmatrix}, x_2 = \begin{pmatrix}1\\0\end{pmatrix}, x_3 = \begin{pmatrix}0\\2\end{pmatrix}, x_4 = \begin{pmatrix}0\\2.5\end{pmatrix}, x_5 = \begin{pmatrix}0\\-1\end{pmatrix}$$

Please list the nested clusters by using hierarchical clustering based on MIN and MAX methods for inter-cluster Euclidean distance measure respectively. Please list all the intermediate Euclidean distance matrices used in the hierarchical clustering process for each method.

(10 marks)

(b) What are the limitations of k-means clustering algorithm? List possible solutions to these limitations.

(5 marks)

(c) Given $N$ data points $x_1, \ldots, x_N$. We aim to build a Gaussian Mixture Model (GMM) with $K$ components to model their distribution. Suppose $\pi_i = P(\omega_i)$ is the prior probability for the $i$th component $\omega_i$, $\mu_i$ is the mean for the $i$th component, $\sigma_i$ is the variance for the $i$th component. Please derive the expression of $\pi_i, \mu_i, \sigma_i$.

(10 marks)

Q4. (25 marks) Face recognition and detection are long-standing and challenging problems in pattern recognition. In this question, you are required to design a face detection and recognition system. Given a collection of face photos from 10 different persons, and each person face is captured by 5 photos of size 30x30. These photos will be used for model training. Please build a face recognition and detection system that includes the following functions: face feature extraction, face classification and face detection, by answering following questions.

(a) How to use PCA to reduce dimensionality of the photo (after vectorization) from 900 to 100 without too much information loss? Please list the computation steps of PCA in details.

(5 marks)

(b) How to use LDA to further reduce dimensionality of the feature (in the above step) from 100 to 8 without losing category information? Please list the computation steps. Is it possible to obtain higher dimensionality (e.g. 20) by LDA? Why?

(5 marks)

(c) For classification, SVM usually gives much better performance than other classification models. Write the formulation of SVM model. Explain why SVM is better than other classification models (e.g., k-nearest neighbor). Explain how to train and test SVM for recognizing faces using the provided photos after dimensionality reduction (PCA and LDA).

(8 marks)

(d) Boosting is a popular detection model due to its high efficiency. Explain the principle behind boosting. Specify the formulation and computation steps for training a detection model using boosting. How to apply this model for face detection on new photos?

(7 marks)

**END OF PAPER**